



Penerapan Algoritma *K-Means* Pada Penjualan Saldo Transportasi Online Studi Kasus Konter XYZ

Hilman Mutaqin

Jurnal Algoritma
STMIK LIKMI

Jl. Ir. H. Juanda No.96, Lebakgede, Kecamatan Cobleng, Kota Bandung, Jawa Barat 40132

Email : info@likmi.ac.id

hilmanm672@gmail.com

Abstrak – Konter XYZ merupakan salah satu konter yang berada di Tasikmalaya dengan salah satu produk yang dijualnya adalah saldo transportasi online seperti Maxim, Gojek, dan Grab. Dengan semakin bertambahnya jumlah pengemudi *online* maka semakin banyak transaksi yang berlangsung sehingga data yang dibuat juga semakin banyak. Akan tetapi data tersebut tidak diolah sehingga semakin menumpuk dan tidak memberikan pengetahuan baru bagi pemilik konter. Tujuan penelitian ini adalah membuat klusterisasi penjualan saldo transportasi online agar menjadi masukan khususnya strategi promosi konter XYZ. Metode yang digunakan adalah *clustering* dengan algoritma *K-Means* agar menghasilkan *cluster* berdasarkan dengan nominal harga pembelian. Dataset yang digunakan dalam penelitian ini berjumlah 4.867 data tetapi setelah di *cleansing* menjadi 2.642 data. Dataset diperoleh dari penjualan konter XYZ pada bulan Oktober tahun 2022. Penelitian ini menghasilkan jumlah *cluster* sebanyak 2 buah dengan *cluster* C1 yaitu nominal rendah sebanyak 2.499 data dengan persentase 94,59%, sedangkan jumlah *cluster* C2 nominal tinggi sebanyak 143 data dengan persentase 5,41. Hasil dari penelitian ini menunjukkan bahwa algoritma *k-means* dapat digunakan untuk *clustering* penjualan saldo transportasi *online* di konter XYZ, serta penjualan saldo transportasi online cukup baik karena jumlah data *cluster* nominal rendah lebih banyak daripada *cluster* nominal tinggi.

Kata Kunci – *Clustering*; *K-Means*; Transaksi; Transportasi *Online*.

I. PENDAHULUAN

Kemajuan teknologi yang pesat dapat membawa perubahan pada hampir semua aspek kehidupan manusia menjadi lebih modern [1]. Salah satu aspek yang berubah adalah layanan transportasi, jika dulu hanya ada layanan transportasi konvensional namun sekarang ada layanan transportasi berbasis aplikasi *online* [2]. Transportasi *online* mempunyai beberapa keunggulan mulai dari aspek keamanan yang diakui konsumen, tarif lebih transparan, dan menyediakan berbagai jenis layanan [3]. Penggunaan transportasi *online* yang semakin meningkat akan membuat jumlah transaksi antara pengemudi *online* dan konter menjadi lebih banyak. Konter akan mempunyai data transaksi dengan jumlah yang banyak. Transaksi merupakan pertukaran barang maupun jasa yang dilakukan secara individu atau organisasi [4]. Konter XYZ adalah salah satu konter yang berada di Tasikmalaya yang menjual produk saldo digital seperti pulsa, voucher internet, uang digital, juga menjual saldo transportasi *online* seperti Maxim, Gojek, Grab, dan lain-lain. Hampir setiap hari transaksi konter XYZ mencapai ratusan sehingga menghasilkan ribuan data transaksi saldo digital perbulan yang disimpan dalam file Microsoft Excel. Akan tetapi data-data tersebut tidak diolah sehingga tidak menghasilkan pengetahuan yang baru. Data yang sebelumnya hanya tertumpuk serta tidak ada pengolahan lebih lanjut tidak akan memberi informasi [5], jika data tersebut diolah maka akan menghasilkan sebuah informasi [6].

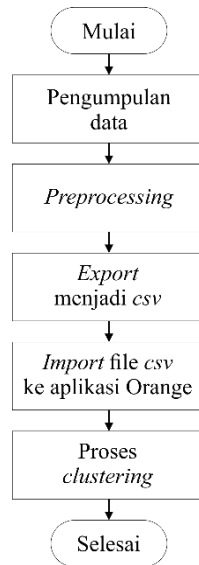
Data mining adalah suatu kumpulan dari beberapa teknik yang digunakan untuk mencari pengetahuan dari

basis data [7]. Data mining juga dipakai untuk mendapatkan informasi baru dan berguna dari jumlah data yang besar [8]. *Clustering* merupakan salah satu dari teknik data mining yang mengelompokkan data berdasarkan kesamaan-kesamaan yang ada dalam sebuah data [9]. Ada banyak jenis model yang dibangun untuk analisis *clustering*, dimana model yang berbeda akan bekerja jauh lebih baik bergantung dari pada perbedaan skenario serta tipe data [10]. *K-Means* adalah salah satu dari banyak algoritma dalam data mining yang masuk kedalam kategori *unsupervised learning*. Dipakai untuk mengelompokkan jenis data yang mempunyai kemiripan kedalam sebuah cluster [11]. *K-means* adalah algoritma yang mudah dan efektif untuk menemukan *cluster* yang ada di dalam data [12]. *K-Means* adalah algoritma sederhana dan umum digunakan [13]. *Silhouette* merupakan sebuah metode untuk mencari jumlah cluster secara ideal serta cepat. Nilai *silhouette* diantara 1 sampai -1, dimana jika semakin tinggi nilainya menandakan bahwa objek tersebut lebih cocok dengan *cluster*-nya sendiri dan tidak cocok dengan cluster lainnya [14]. Orange merupakan aplikasi kode terbuka penambang data. Aplikasi Orange dapat dipakai untuk melakukan analisis dan visualisasi data. Aplikasi Orange disukai karena faktor inovatif, kualitas dan kendala. Aplikasi Orange dapat mempermudah penggunaannya untuk melakukan eksperimen dengan data serta proses data analisis secara visual dan intuitif [15].

Penelitian terdahulu dengan metode *clustering* serta algoritma *K-Means* yang pertama dengan tema sebaran pelanggan serta karakteristiknya menghasilkan 3 *cluster* yaitu *cluster* C1 3 kelompok pelanggan dengan transaksi banyak dan bertempat tinggal di Malang Kota, C2 3 kelompok pelanggan dengan transaksi sedang yang bertempat tinggal di Malang Kota dan Kabupaten, dan C3 2 kelompok pelanggan dengan transaksi sedikit serta bertempat tinggal diluar Malang [16]. Penelitian kedua dengan tema mengelompokkan harga beras di Indonesia menghasilkan 3 *cluster* yaitu *cluster* harga tinggi sebanyak 11 kota, cluster harga sedang sebanyak 11 kota dan cluster harga rendah sebanyak 11 kota [17]. Penelitian ketiga dengan tema pemetaan pasar berdasarkan harga pangan menghasilkan 3 *cluster* yaitu *cluster* tinggi dengan jumlah 1 pasar, *cluster* rendah dengan jumlah 10 pasar, dan cluster sedang berjumlah 17 pasar [18]. Penelitian keempat dengan tema pengelompokan harga cabai rawit menghasilkan 3 *cluster* yaitu *cluster* dengan tingkat harga tinggi sebanyak 10 provinsi, *cluster* dengan harga rendah sebanyak 12 provinsi dan dengan harga sedang sebanyak 12 provinsi [19]. Penelitian terakhir dengan tema pengelompokan penjualan aksesoris menghasilkan 3 *cluster* yaitu cluster kurang laku sebanyak 12 data, *cluster* laku sebanyak 10 data, dan *cluster* paling laku sebanyak 4 data [20]. Penelitian sebelumnya menggunakan jumlah data yang sedikit sedangkan penelitian ini menggunakan jumlah data mencapai 4.867 data. Penelitian ini menggunakan pendekatan berbeda untuk mencari jumlah *cluster* yaitu *silhouette* sedangkan penelitian sebelumnya menggunakan pendekatan *elbow*. Aplikasi yang digunakan adalah Orange serta yang akan dianalisis adalah nominal pembelian saldo transportasi online. Tujuan dari penelitian ini untuk melakukan klasterisasi penjualan saldo Maxim, Gojek dan Grab berdasarkan nominal dan Trx agar didapatkan informasi yang berguna bagi pemilik konter serta membantu dalam pengambilan keputusan seperti strategi promosi maupun penentuan stok saldo.

II. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan menggunakan data sekunder. Tahapan penelitian dijelaskan pada gambar 1.



Gambar 1: Tahapan Penelitian

Penelitian ini dimulai dengan melakukan pengumpulan data yaitu melakukan wawancara dan mengumpulkan data penjualan pada bulan Oktober tahun 2022 berjumlah 4.876 data dari pemilik konter XYZ. Data yang telah diperoleh terlebih dahulu di *cleaning* seperti menghilangkan transaksi selain angkutan online, kolom No Hp dihapus, dan kolom tanggal dirubah menjadi hari. Data yang telah di *cleaning* selanjutnya di *export* menjadi file *csv* dan di *import* ke dalam aplikasi Orange. Data yang telah di *import* ke dalam aplikasi akan diolah untuk mencari *cluster* dari data tersebut. proses *clustering* dilakukan secara otomatis dengan aplikasi. Tahap terakhir dalam penelitian ini adalah melakukan analisis dari proses *cluster* yang telah dibuat untuk mendapatkan informasi baru dari data yang telah diproses.

III. HASIL DAN DISKUSI

1. Pengumpulan data

Data yang diperoleh merupakan data hasil penjualan dari konter XYZ pada bulan Oktober tahun 2022. Jumlah dataset yang diperoleh sebanyak 4.876 data. Tabel 1 merupakan dataset penjualan konter XYZ.

Tabel 1: Dataset penjualan konter XYZ

No	Tanggal	Trx	No Hp	Harga
1	44835	S25	xxxxxxxxxxxx	27000
2	sabtu	PLN50	xxxxxxxxxxxx	52000
3		T10	xxxxxxxxxxxx	12000
4		S15	xxxxxxxxxxxx	17000
5		S15	xxxxxxxxxxxx	17000
6		S5	xxxxxxxxxxxx	7000
7		T10	xxxxxxxxxxxx	12000
8		S10	xxxxxxxxxxxx	12000
.	.	.	xxxxxxxxxxxx	.
.	.	.	xxxxxxxxxxxx	.
4876		m10	xxxxxxxxxxxx	10000

2. Preprocessing

Dataset yang telah diperoleh tidak bisa langsung diproses namun data tersebut perlu di *cleaning* terlebih

dahulu pada aplikasi Microsoft Excel, seperti menghapus transaksi selain transportasi online, memperbaiki kolom hari, merubah kolom Trx dengan jenis transportasi online dan menghapus no hp agar bisa diproses menggunakan aplikasi Orange. Setelah data di *cleaning* hanya ada 2.642 data penjualan transportasi online. Tabel 2 merupakan dataset penjualan konter XYZ yang telah di *cleaning*.

Tabel 2: Dataset yang telah di *cleaning*

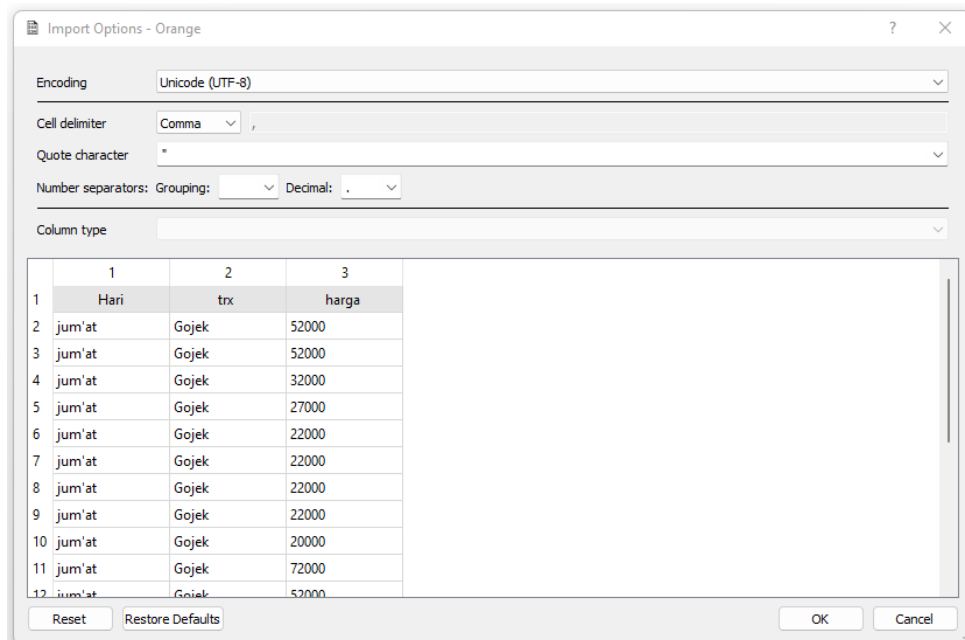
No	Hari	Trx	Harga
1	Jum'at	Gojek	52000
2	Jum'at	Gojek	52000
3	Jum'at	Gojek	32000
4	Jum'at	Gojek	27000
5	Jum'at	Gojek	22000
6	Jum'at	Gojek	22000
7	Jum'at	Gojek	22000
8	Jum'at	Gojek	22000
.	.	.	.
2642	Senin	Maxim	23000

3. *Export* menjadi csv

Setelah data di *cleaning*, selanjutnya adalah *export* file tersebut dengan format *csv* agar dapat diproses dengan aplikasi orange.

4. *Import* file *csv* ke aplikasi Orange

Data yang telah di *export* bisa langsung di *import* ke dalam aplikasi Orange. Gambar 2 adalah proses *import* file *csv* kedalam aplikasi Orange.

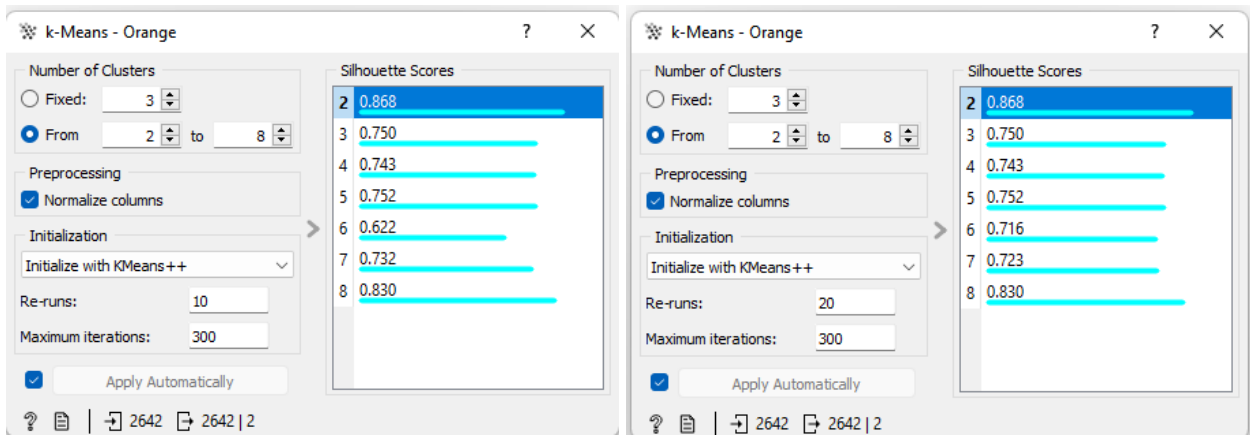


Gambar 2: proses *import* file *csv* kedalam aplikasi Orange

5. Proses *clustering*

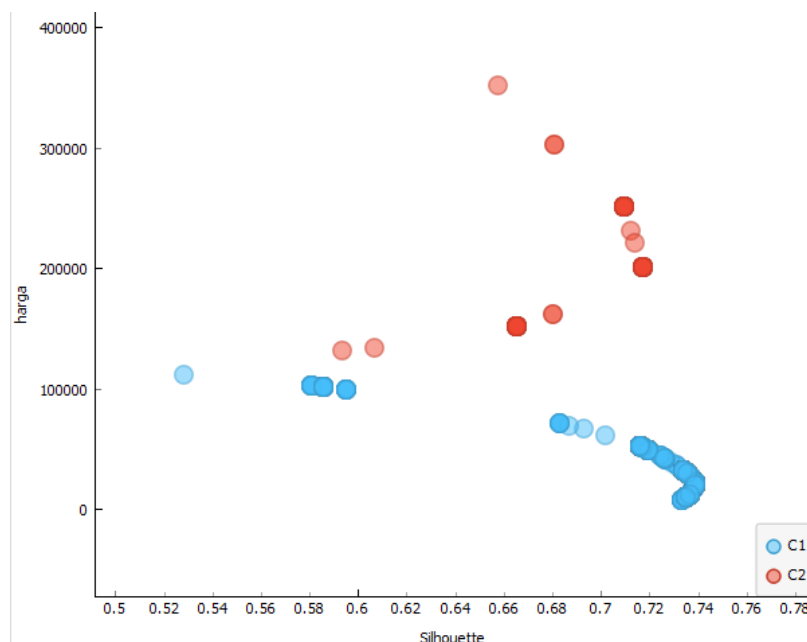
Pembuatan jumlah *cluster* dibuat otomatis oleh aplikasi, akan tetapi perlu dilakukan pengaturan terlebih dahulu pada fitur *k-means* untuk memperoleh jumlah *cluster* yang paling ideal. *Number of Cluster* diatur dari 2 sampai 8 *cluster*. Algoritma dijalankan sebanyak 2 kali yaitu 10 dan 20 kali pengulangan agar hasil

yang didapat lebih konsisten. Setelah fitur *k-means* diatur maka diperoleh jumlah *cluster* yang ideal yaitu 2 *cluster* dengan *silhouette Score* tertinggi yaitu 0,868. Proses pemilihan jumlah *cluster* ditampilkan dalam gambar 3.



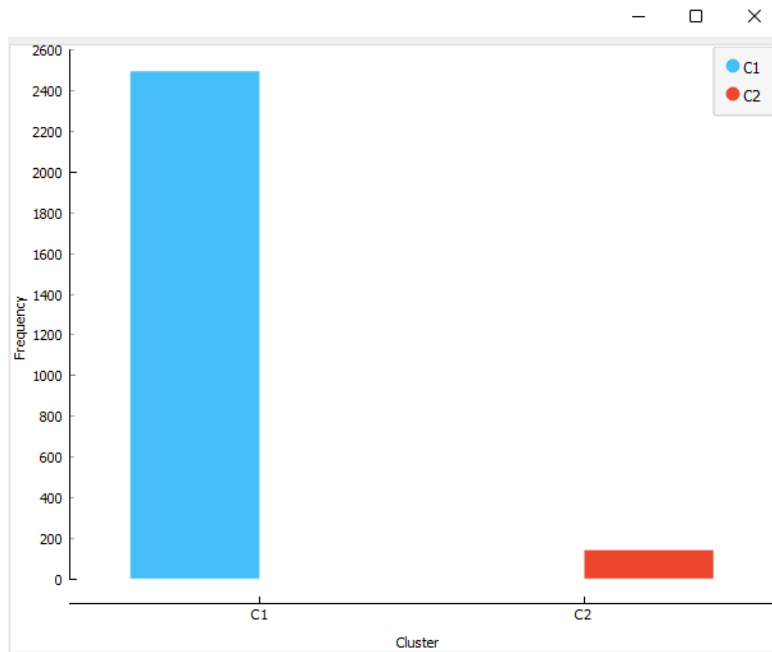
Gambar 3: Proses pemilihan jumlah *cluster* dengan aplikasi Orange

Penyebaran file dalam *cluster* dibuat oleh aplikasi. Nilai terendah pada *cluster* C1 yaitu nominal rendah adalah 8.000 sedangkan nilai tertinggi tertinggi adalah 112.000, sedangkan nilai terendah *cluster* C2 yaitu nominal tinggi adalah 132.000 sedangkan nilai tertinggi adalah 353.000. Dari hasil tersebut diketahui bahwa keuntungan maksimal bisa didapat jika pembeli melakukan *top up* saldo dibawah 112.000. Gambar 4 merupakan tampilan penyebaran data dengan menggunakan Scatter Plot.



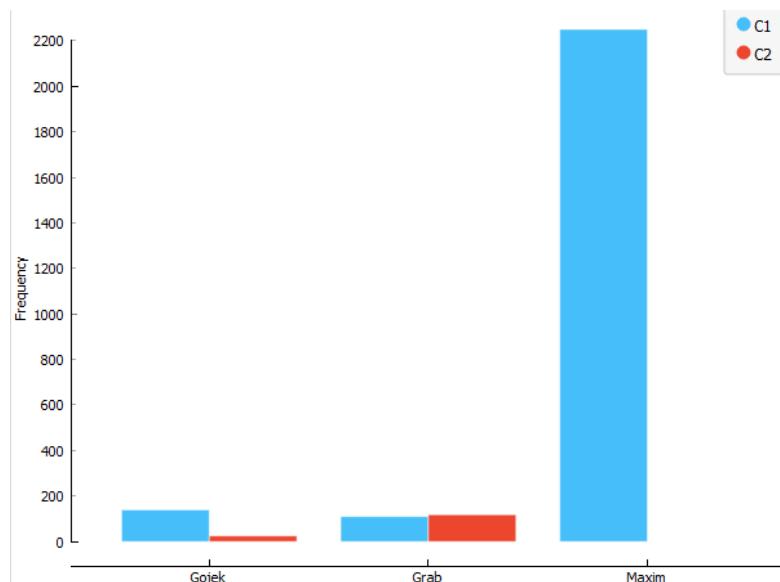
Gambar 4: Sebaran data dengan aplikasi Orange

Jumlah dan presentasi *cluster* digambarkan dengan diagram batang. Jumlah *cluster* C1 sebanyak 2.499 data dengan persentase 94,59%, sedangkan jumlah *cluster* C2 sebanyak 143 data dengan persentase 5,41%. Dari hasil tersebut diketahui bahwa penjualan saldo transportasi online konter XYZ cukup baik karena lebih banyak data pada *cluster* C1. Gambar 5 merupakan tampilan diagram batang berdasarkan cluster dari file yang telah diproses.



Gambar 5: Diagram batang jumlah data cluster C1 dan cluster C2

Dari Gambar 6 bisa dilihat bahwa jumlah data Gojek adalah 139 data pada *cluster* C1 dan 25 data pada *cluster* C2, sedangkan Grab adalah 110 data pada *cluster* C1 dan 118 data pada *cluster* C2, dan Maxim sebanyak 2.250 data pada *cluster* C1 dan tidak ada data pada *cluster* C2. Dari hasil tersebut diketahui bahwa penjualan Maxim sudah memberikan keuntungan yang maksimal karena semua data berada pada C1 sedangkan Gojek dan Grab belum memberikan keuntungan maksimal karena masih ada data yang berada pada *cluster* C2 dengan paling banyak adalah Grab. Gambar 6 menampilkan diagram batang berdasarkan Trx.



Gambar 6: Diagram batang jumlah berdasarkan Trx

Hasil dari penelitian ini menunjukkan bahwa algoritma *k-means* dapat digunakan untuk *clustering* penjualan saldo transportasi online di konter XYZ, dan penjualan saldo transportasi online cukup baik karena jumlah data cluster nominal rendah lebih banyak daripada cluster nominal tinggi. Akan tetapi ada

beberapa data yang berada pada *cluster* C2 sehingga belum memberikan keuntungan maksimal bagi pemilik konter. Pada penelitian sebelumnya semua *cluster* yang dihasilkan berjumlah 3 dengan dalah satu metode yang digunakan untuk mencari jumlah *cluster* adalah *elbow*, sedangkan pada penelitian ini dengan algoritma *k-means* dan metode *silhouette* serta menggunakan aplikasi Orange, jumlah *cluster* yang dihasilkan adalah 2.

IV. KESIMPULAN

Dari penelitian yang telah dilakukan, maka dapat disimpulkan bahwa *cluster* yang dibuat sebanyak 2 buah dengan jumlah *cluster* C1 yaitu nominal rendah sebanyak 2.499 data dengan persentase 94,59%, sedangkan jumlah *cluster* C2 yaitu nominal tinggi sebanyak 143 data dengan persentase 5,41%. Sedangkan berdasarkan Trx adalah jumlah data Gojek adalah 139 data pada *cluster* C1 dan 25 data pada *cluster* C2, sedangkan Grab adalah 110 data pada *cluster* C1 dan 118 data pada *cluster* C2, dan Maxim sebanyak 2.250 data pada *cluster* C1 dan tidak ada data pada *cluster* C2. Menurut pemilik konter, semakin kecil nominal pembelian maka semakin banyak jumlah transaksi yang dibuat dengan jumlah modal saldo yang sama sehingga keuntungan bisa maksimal. Data Gojek dan Grab ada pada *cluster* C2 sehingga dengan modal yang sama, jumlah keuntungan tidak maksimal karena jumlah transaksi sedikit. Oleh karena itu pemilik konter sebaiknya membuat promosi agar data pada Gojek dan Grab bisa berada pada *cluster* C1 sehingga jumlah transaksi yang dibuat menjadi lebih banyak dengan modal saldo yang sama yang pada akhirnya keuntungan bisa lebih maksimal. Saran untuk penelitian selanjutnya yaitu data yang digunakan lebih banyak agar hasil yang didapat bisa lebih mempresentasikan hasil penjualan saldo transportasi online secara umum.

DAFTAR PUSTAKA

- [1] N. Huda, "Analisis Kinerja Website PT PLN (Persero) Menggunakan Metode Pieces," *Sistemasi*, vol. 8, no. 1, pp. 78–89, 2019, doi: DOI: <https://doi.org/10.32520/stmsi.v8i1>.
- [2] P. Fakhriyah, "Pengaruh Layanan Transportasi Online (Gojek) Terhadap Perluasan Lapangan Kerja Bagi Masyarakat Di Kota Cimahi," *Comm-Edu (Community Educ. Journal)*, vol. 3, no. 1, p. 34, 2020, doi: [10.22460/comm-edu.v3i1.3719](https://doi.org/10.22460/comm-edu.v3i1.3719).
- [3] K. Setiyorini and G. Hendrastomo, "Persaingan Antara Ojek Online Dengan Ojek Konvensional Di Stasiun Lempuyangan, Daerah Istimewa Yogyakarta," *J. Sosiol. Pendidik. Humanis*, vol. 3, no. 1, p. 29, 2019, doi: [10.17977/um021v3i1p29-35](https://doi.org/10.17977/um021v3i1p29-35).
- [4] A. Nasir and Suhendi, "Penerapan Pengelolaan Transaksi Keuangan Menggunakan Modul Accounting And Finance Odoo 10 Studi Kasus Yayasan Sdit Bahrul Fikri," *J. Inform. Terpadu*, vol. 4, no. 1, p. 2, 2018, doi: <https://doi.org/10.54914/jit.v4i1.120>.
- [5] Nawassyarif, M. Julkarnain, and K. Rizki Ananda, "Sistem Informasi Pengolahan Data Ternak Unit Pelaksana Teknis Produksi Dan Kesehatan Hewan Berbasis Web," *J. Inform. Teknol. dan Sains*, vol. 2, no. 1, pp. 32–39, 2020, doi: [10.51401/jinteks.v2i1.556](https://doi.org/10.51401/jinteks.v2i1.556).
- [6] A. Syafnur, "Analisis Dengan Metode Klasifikasi Menggunakan Decission Tree Untuk Memprediksi Penentuan Resiko kredit Bank," *Jurteks*, vol. 4, no. 1, pp. 101–106, 2017, doi: [10.33330/jurteks.v4i1.30](https://doi.org/10.33330/jurteks.v4i1.30).
- [7] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, Apr. 2021, doi: [10.29207/resti.v5i2.3007](https://doi.org/10.29207/resti.v5i2.3007).
- [8] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 187–192, 2021, doi: [10.29207/resti.v5i1.2813](https://doi.org/10.29207/resti.v5i1.2813).
- [9] T. Syahputra, J. Halim, and E. P. Sintho, "Penerapan Data Mining Dalam Menentukan Pilihan Jurusan Bidang Studi SMA Menggunakan Metode," *Jurteks*, vol. 4, no. 2, pp. 1–4, 2018, doi: <https://doi.org/10.33330/jurteks.v4i2.61>.
- [10] C. Anggarwal, *Data Mining*, vol. 14, no. 3. New York: Springer, 2015.
- [11] R. Ananda, M. Zidny Naf'an, A. B. Arifa, and A. Burhanuddin, "Sistem Rekomendasi Pemilihan

- Peminatan Menggunakan Density Canopy K-Means,” *RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 172–179, 2020, doi: <https://doi.org/10.29207/resti.v4i1.1531>.
- [12] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: an Introduction to Data Mining*, vol. 100, no. 472. New Jersey: John Wiley & Sons, 2005.
- [13] N. Agustina and P. Prihandoko, “Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 3, pp. 621–626, 2018, doi: [10.29207/resti.v2i3.492](https://doi.org/10.29207/resti.v2i3.492).
- [14] A. Punhani, N. Faujdar, K. K. Mishra, and M. Subramanian, “Binning-based Silhouette Approach to Find the Optimal Cluster using K-Means,” *IEEE Access*, vol. 10, no. October, pp. 115025–115032, 2022, doi: [10.1109/ACCESS.2022.3215568](https://doi.org/10.1109/ACCESS.2022.3215568).
- [15] R. A. raffaidy Wiguna and A. I. Rifai, “Analisis Text Clustering Masyarakat Di Twitter Mengenai Omnibus Law Menggunakan Orange Data Mining,” *J. Inf. Syst. Informatics*, vol. 3, no. 1, pp. 1–12, 2021, doi: [10.33557/journalisi.v3i1.78](https://doi.org/10.33557/journalisi.v3i1.78).
- [16] F. Nurdianyansyah, S. Arifin, and F. Marisa, “Penerapan Clustering Algorithm Untuk Mendukung Promosi Server Pulsa Reload,” *J I M P - J. Inform. Merdeka Pasuruan*, vol. 3, no. 2, pp. 73–78, 2018, doi: [10.37438/jimp.v3i2.174](https://doi.org/10.37438/jimp.v3i2.174).
- [17] T. Siburian, M. Safii, and I. Parlina, “Penerapan Algoritma K-Means Clustering untuk Pengelompokan Harga Eceran Beras di Pasar Tradisional Berdasarkan Wilayah Kota,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. 100, pp. 927–936, 2019, doi: [10.30645/senaris.v1i0.101](https://doi.org/10.30645/senaris.v1i0.101).
- [18] A. Novita and H. B. Seta, “Pemetaan Pasar Tradisional Berdasarkan Harga Pangan Komoditas Menggunakan Algoritma K-Means,” *Senamika*, vol. 2, no. 2, pp. 320–326, 2021.
- [19] C. F. Palembang and S. P. Palembang, “Pengelompokan Tingkat Harga Cabai Rawit Berdasarkan Provinsi di Indonesia Menggunakan Algoritma K-Means Clustering,” *Variance*, vol. 3, no. 2, pp. 48–60, 2021, doi: [/10.30598/variancevol3iss2](https://doi.org/10.30598/variancevol3iss2).
- [20] M. Mardalius, “Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means,” *Jurteks*, vol. 4, no. 2, pp. 123–132, 2018, doi: [10.33330/jurteks.v4i2.36](https://doi.org/10.33330/jurteks.v4i2.36).